The Voice of Optimization

Bartolomeo Stellato

joint work with Dimitris Bertsimas

IDSS Seminar on Algebra, Statistics, and Optimization 2019

Inventory management

$$
\begin{array}{ll}
\text{minimize} & \displaystyle\sum_{t=0}^{T-1} h(x_t) + o(u_t) \\
\text{subject to} & \boxed{x_{t+1}} = x_t + \boxed{u_t} - \boxed{d_t} \\
& x_0 = \boxed{x_{\text{init}}} \\
& 0 \leq u_t \leq M
\end{array}
$$

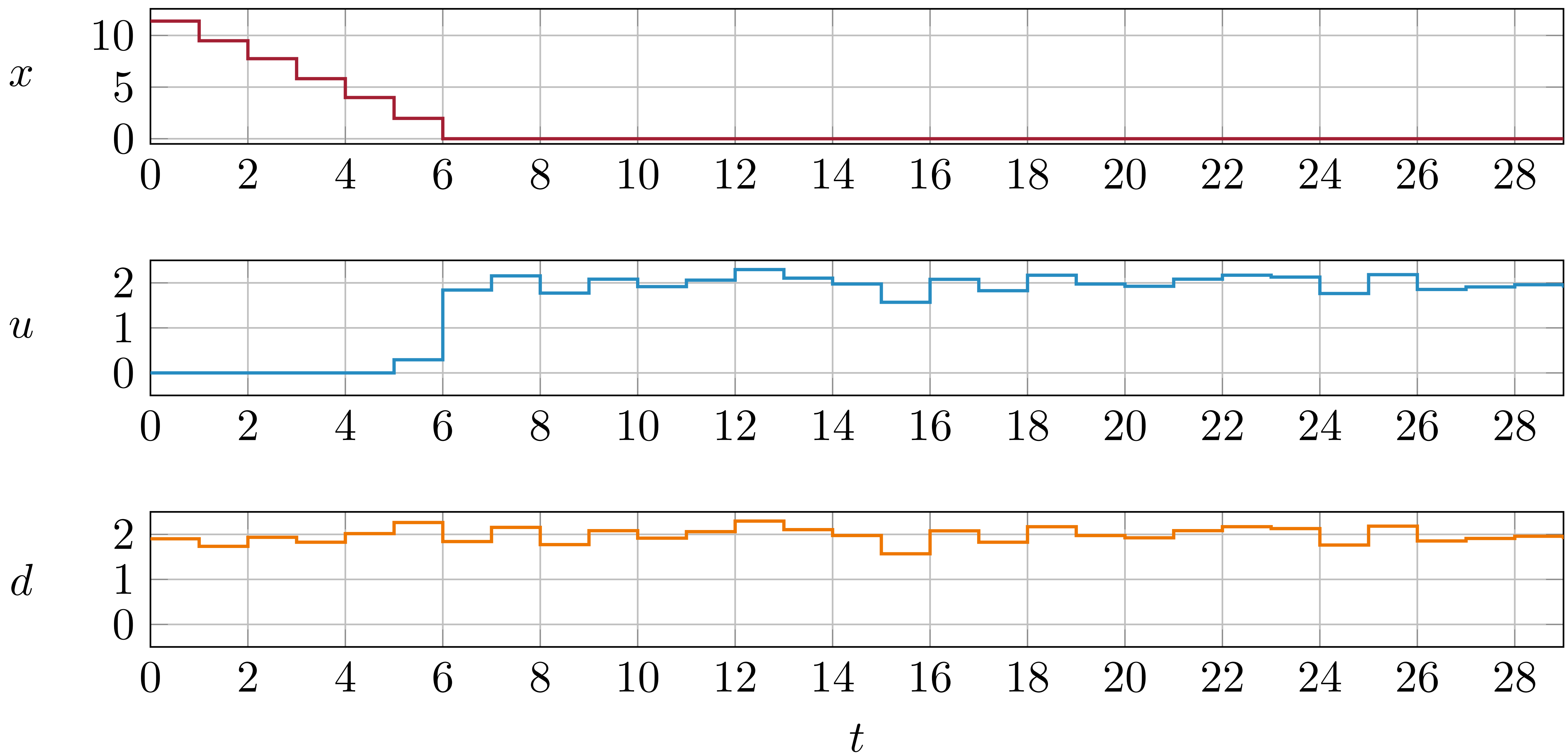
Inventory

Order

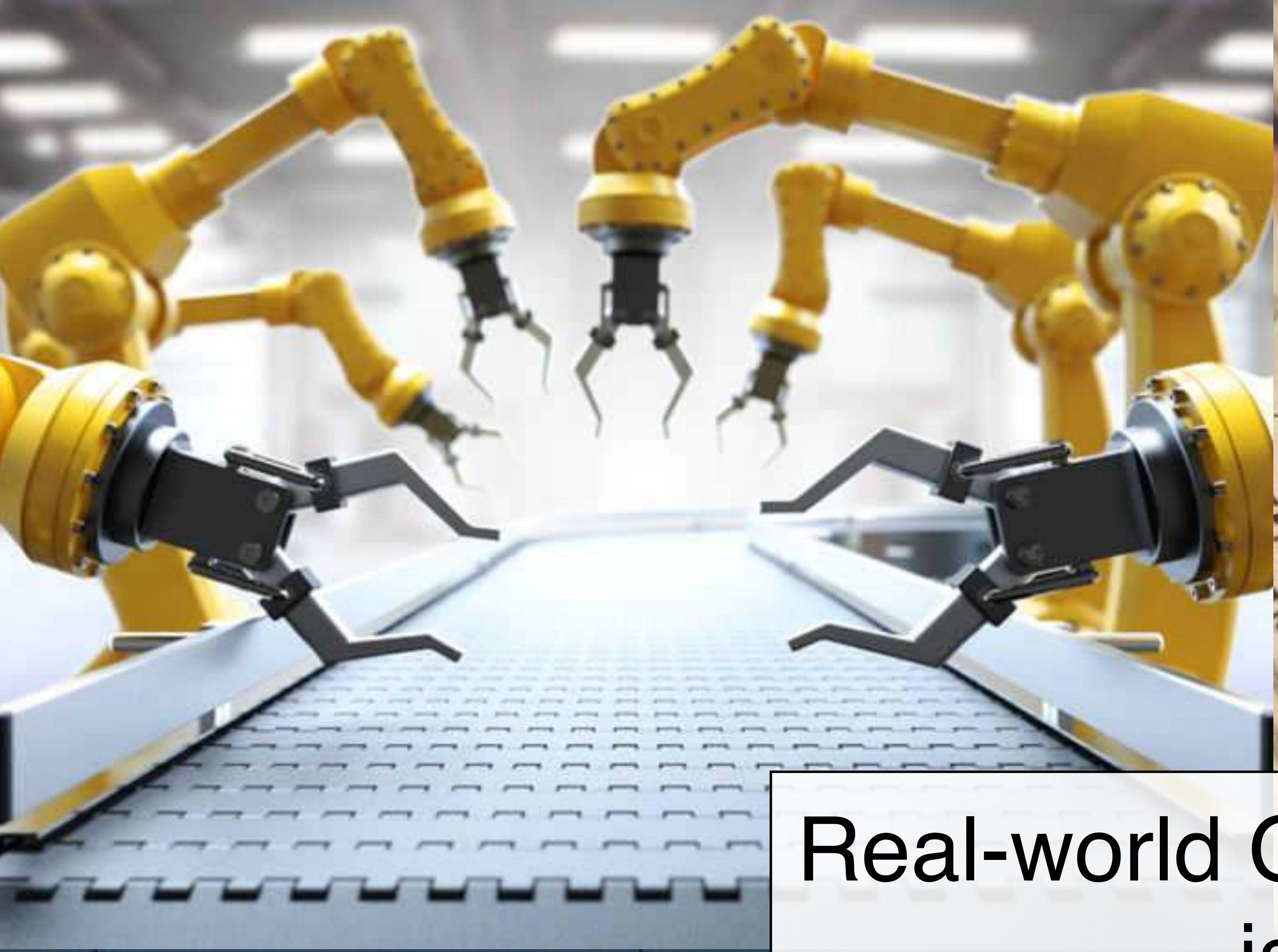
Demand

Parameters

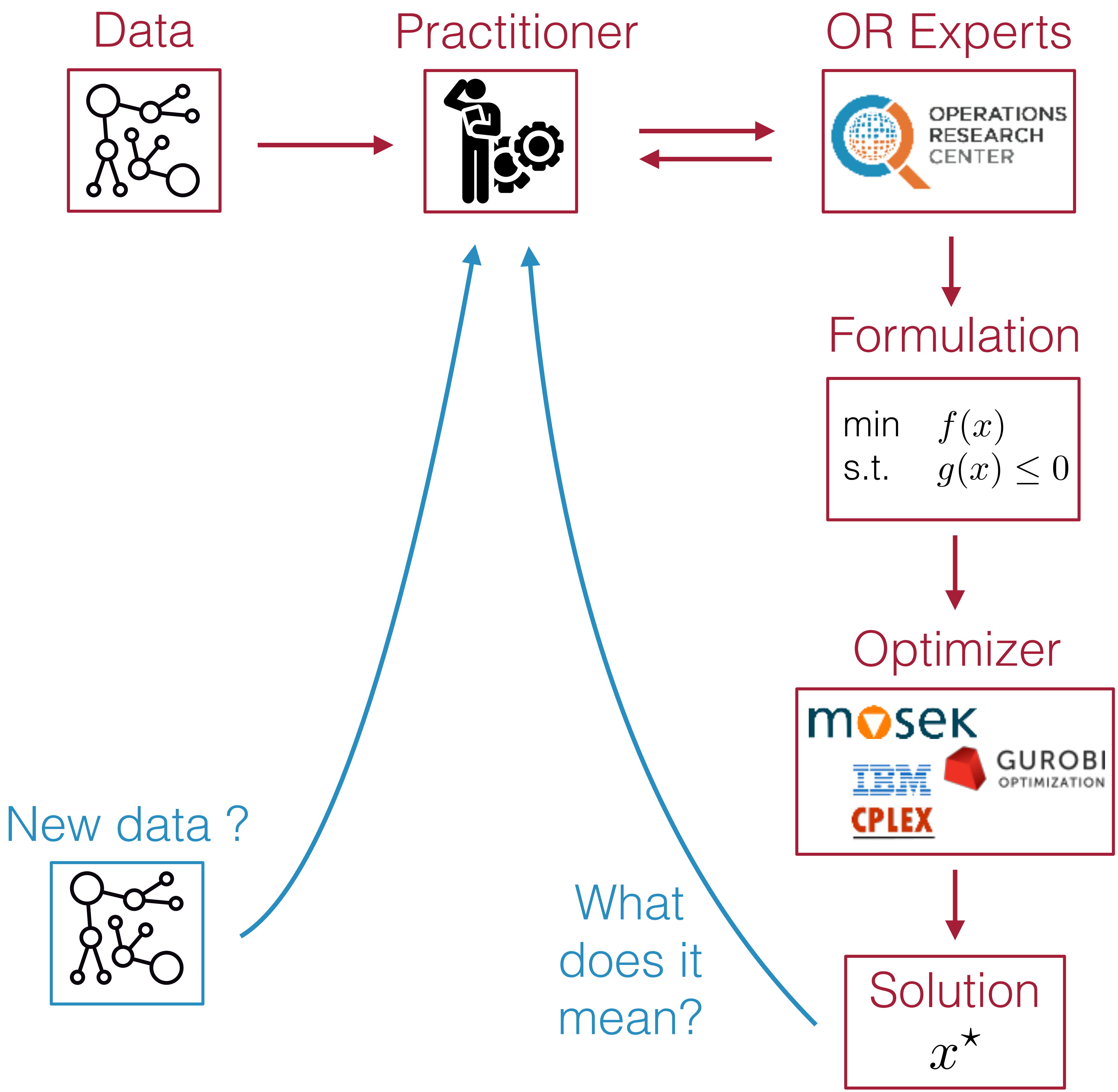
What does it mean?



x_{init} ? d_t ?



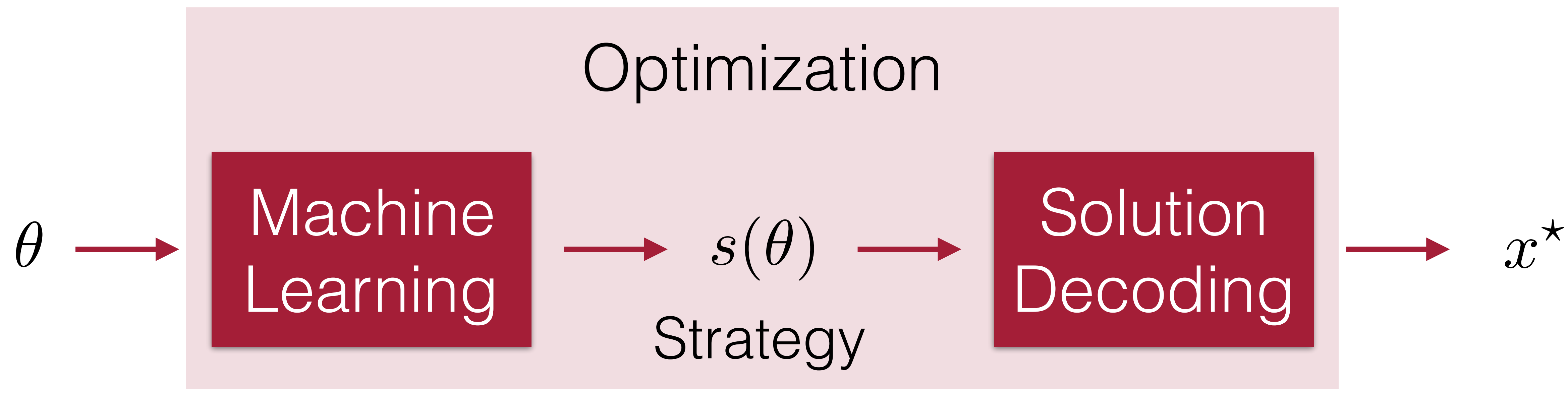
Real-world Optimization
is
Parametric


Data
Practitioner
OR Experts
OPERATIONS RESEARCH CENTER
Formulation
min $f(x)$
s.t. $g(x) \leq 0$
Optimizer
mosek
IBM
CPLEX
GUROBI OPTIMIZATION
Solution
$x^\star$
New data ?
What does it mean?

$\theta \longrightarrow$ $\boxed{\text{Optimization}}$ $\longrightarrow x^\star$

Black box $\longrightarrow$ Open the box

Too slow $\longrightarrow$ Solve in milliseconds



Black box $\longrightarrow$ Open the box

Too slow $\longrightarrow$ Solve in milliseconds

Optimal strategies

Strategy

The complete information to efficiently get the optimal solution

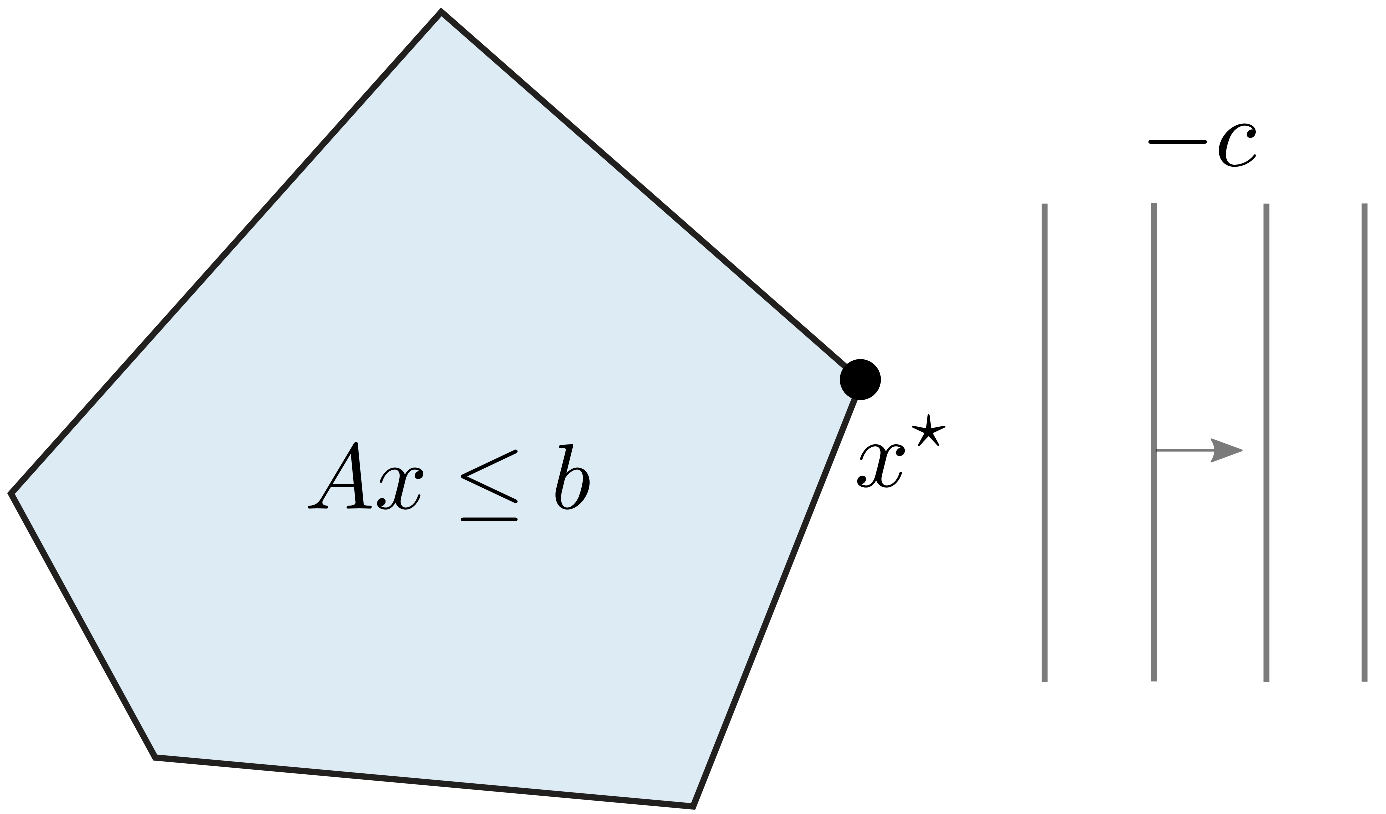
$$s(\theta)$$

Linear optimization

$$\begin{array}{ll} \text{minimize} & c(\theta)^T x \\ \text{subject to} & A(\theta) x \leq b(\theta) \end{array}$$

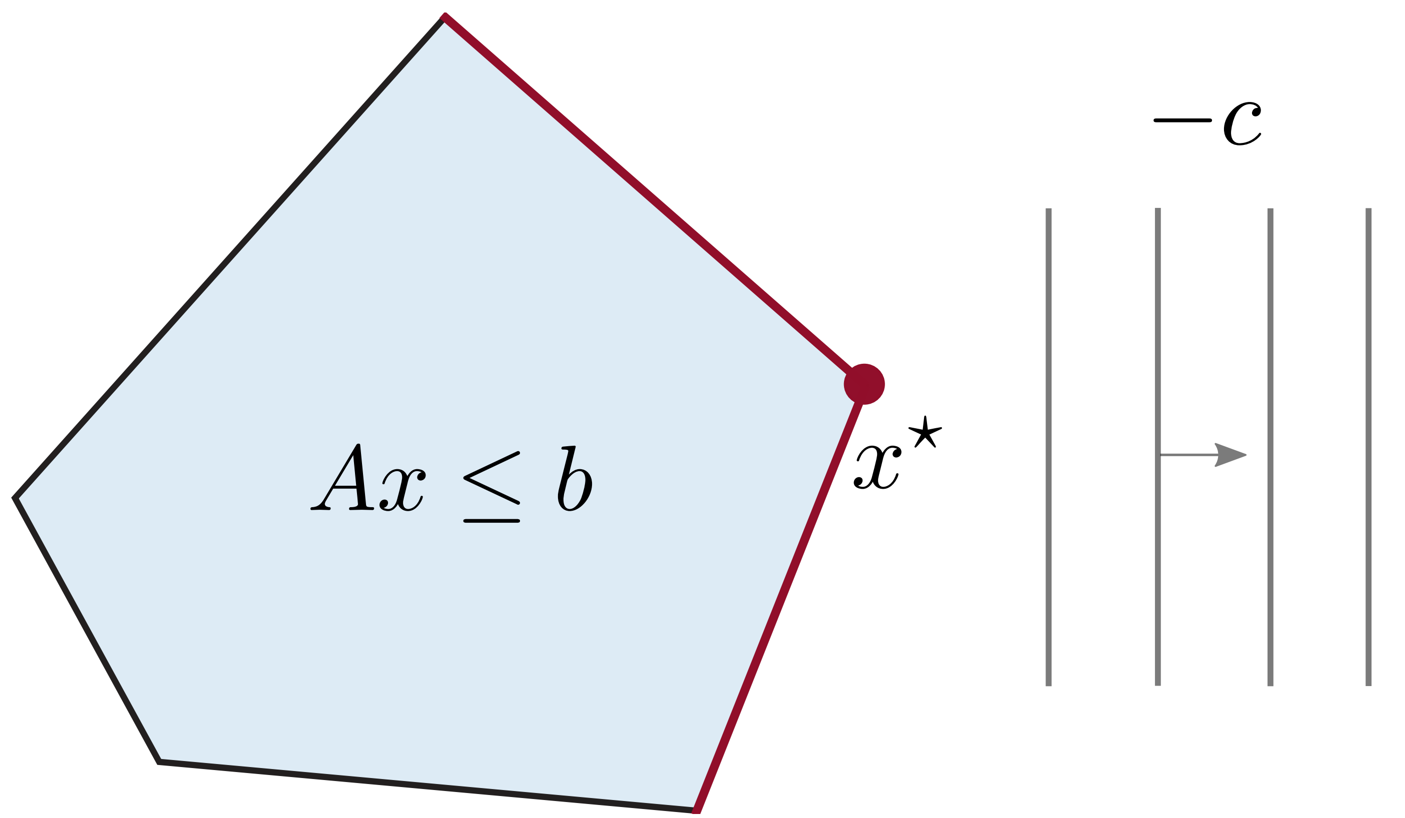
Parameters

Variables



Tight constraints

$$\mathcal{T}(\theta) = \{i \mid A_i(\theta)x^\star = b_i(\theta)\}$$



$|\mathcal{T}(\theta)| = \#$ variables — if non-degenerate

$|\mathcal{T}(\theta)| \ll \#$ constraints — in general

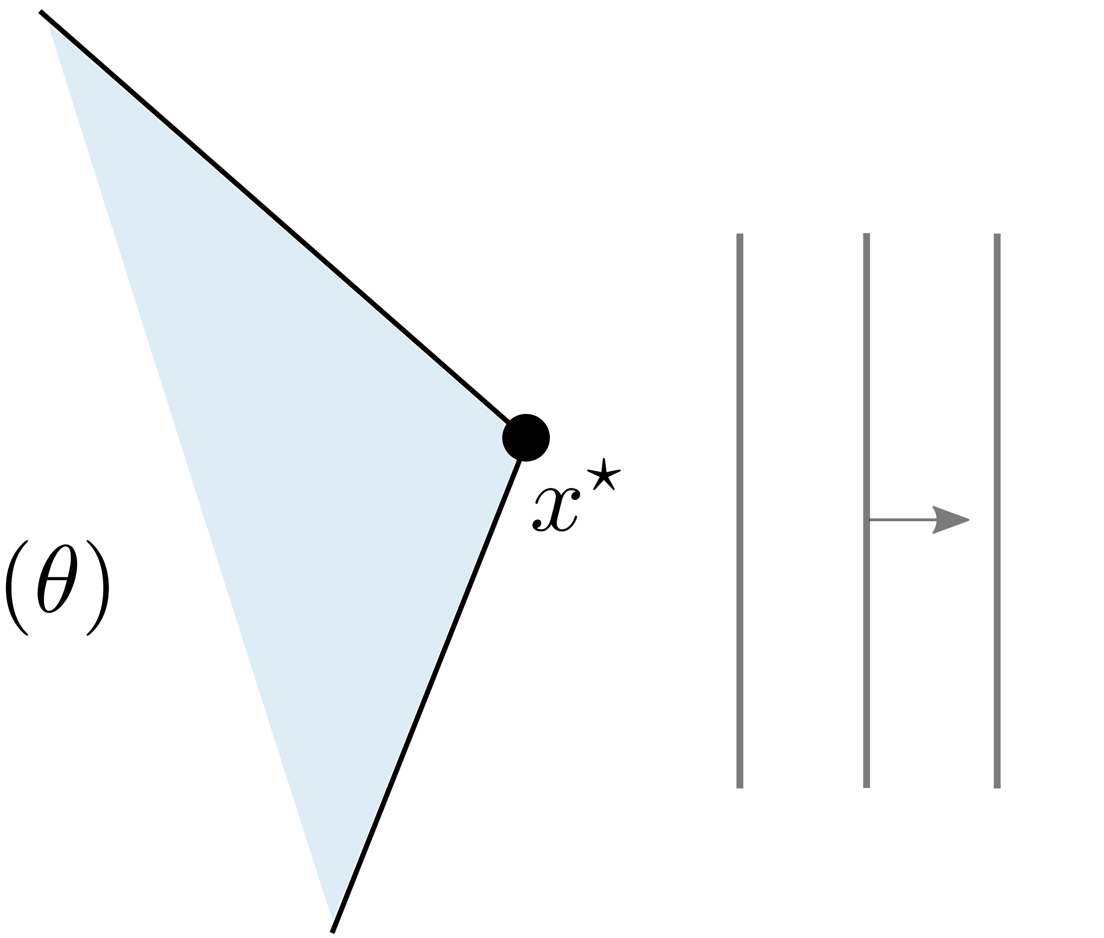
Strategy

The complete information to efficiently get the optimal solution

$$s(\theta) = \mathcal{T}(\theta)$$

$s(\theta) \longrightarrow$ **Solution Decoding** $\longrightarrow x^\star$

$$\begin{array}{ll} \text{minimize} & c(\theta)^T x \\ \text{subject to} & A_i(\theta)x = b_i(\theta) \quad \forall i \in \mathcal{T}(\theta) \end{array}$$



Inventory management

$$
\begin{aligned}
\text{minimize} \quad & \sum_{t=0}^{T-1} h(x_t) + o(u_t) \\
\text{subject to} \quad & x_{t+1} = x_t + u_t - d_t \\
& x_0 = x_{\text{init}} \\
& 0 \leq u_t \leq M
\end{aligned}
$$

Inventory

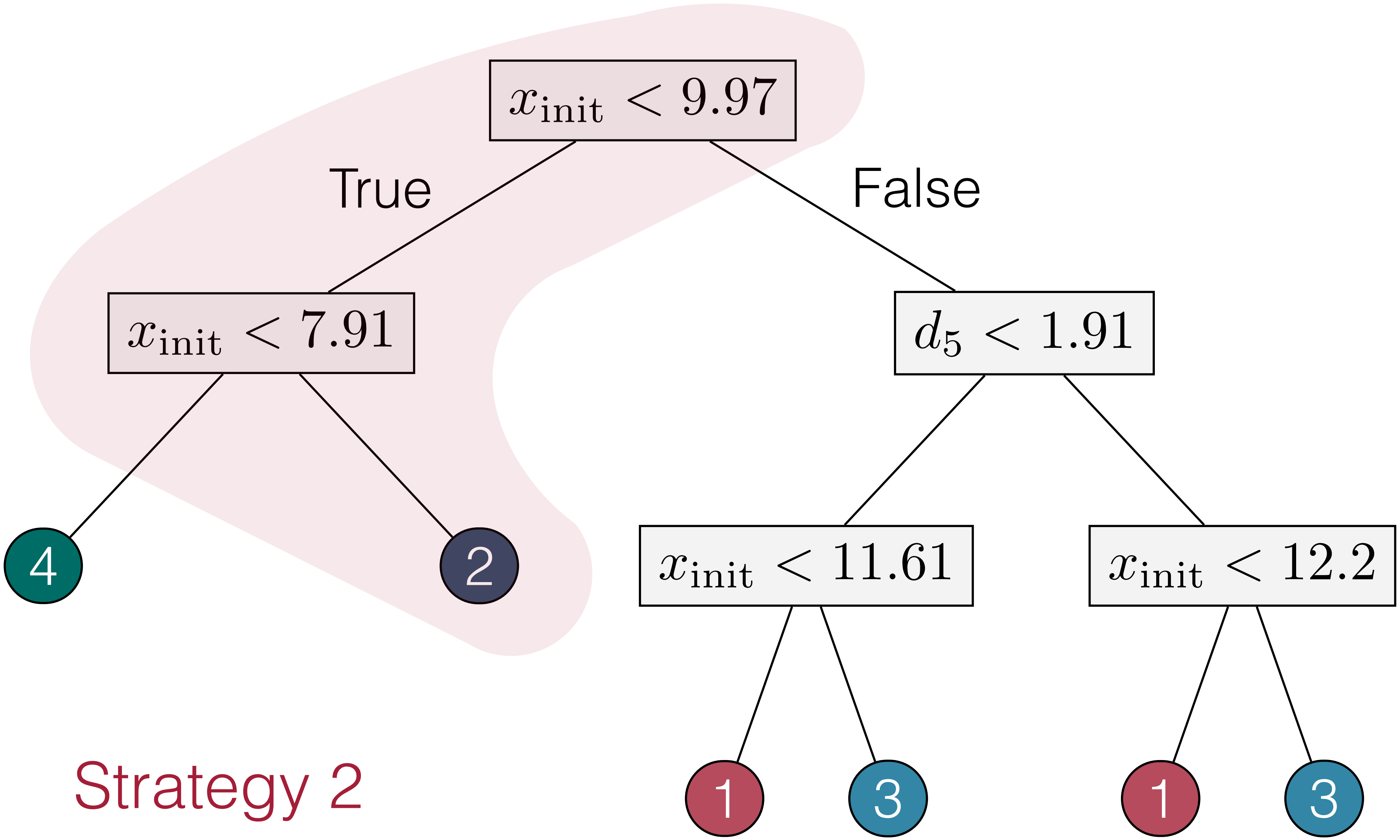
Order

Demand

Inequalities

Parameters

Strategy selection



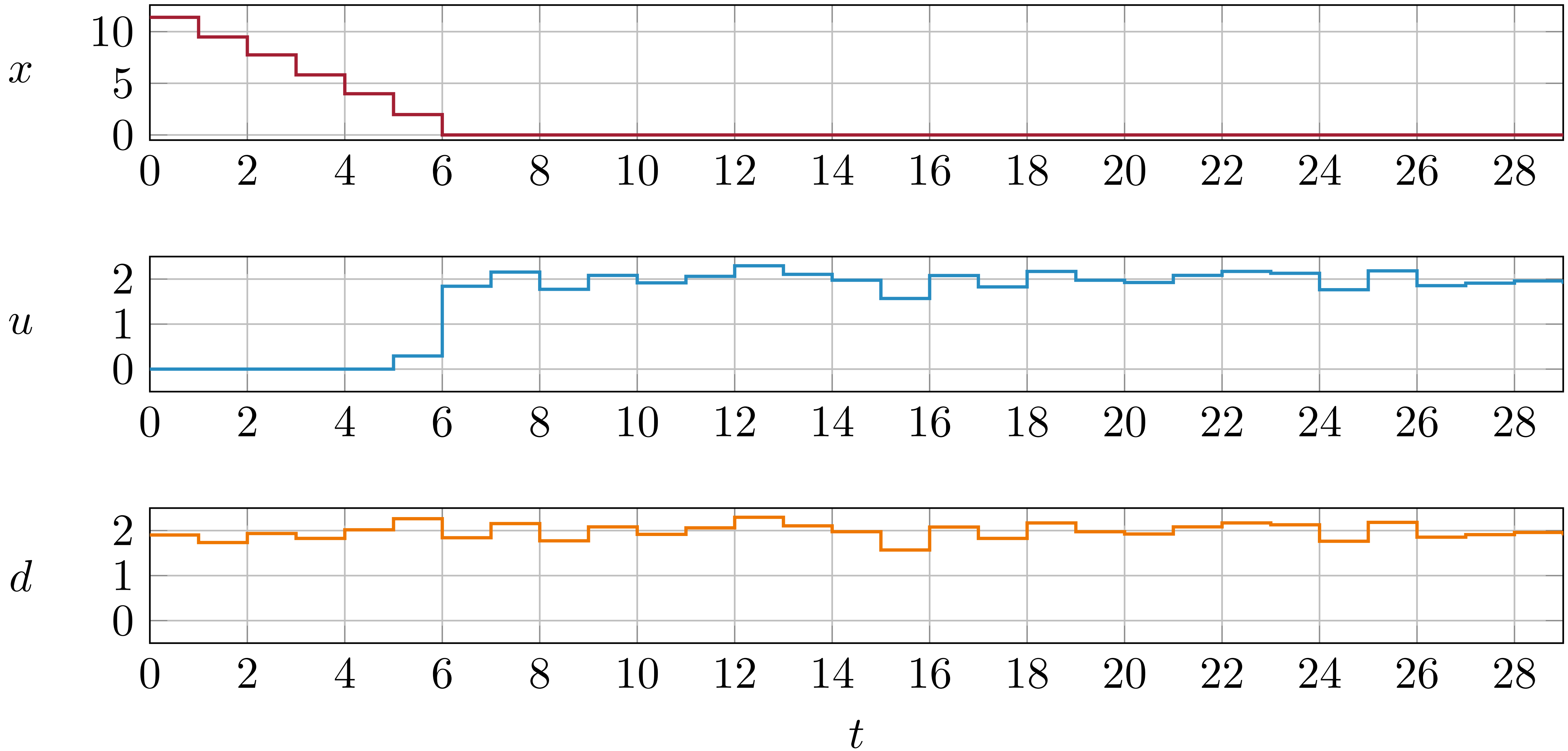
Strategy 2

$$u_t = 0 \qquad t \le 4$$
$$0 \le u_t \le M \qquad t > 4$$

Strategies for inventory

Strategy 2

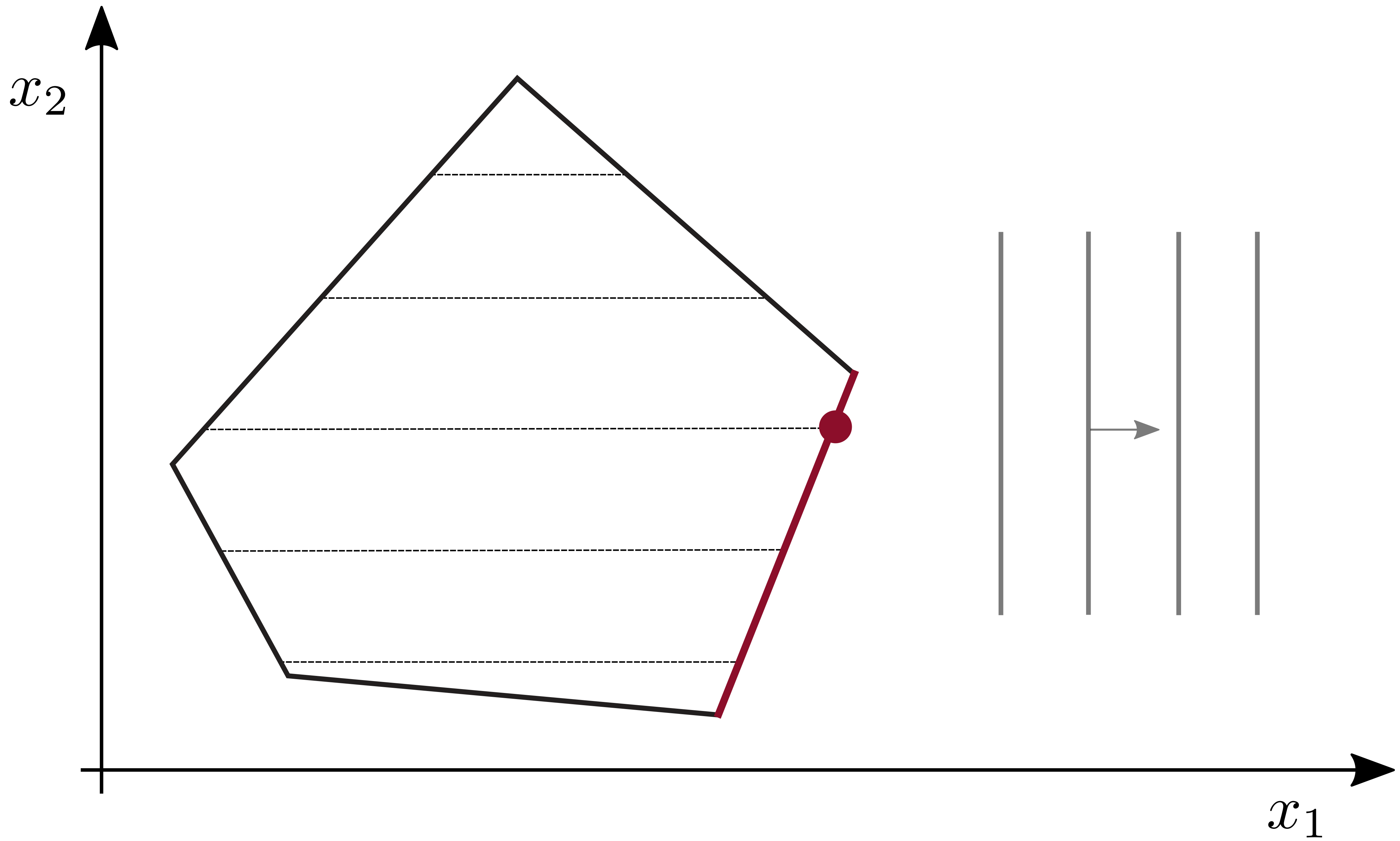
$$\begin{array}{ll} u_t = 0 & t \leq 4 \\ 0 \leq u_t \leq M & t > 4 \end{array}$$



Mixed-integer optimization

$$
\begin{array}{ll}
\text{minimize} & c(\theta)^T x \\
\text{subject to} & A(\theta)x \leq b(\theta) \\
 & \boxed{x_{\mathcal{I}} \in \mathbf{Z}^d} \quad \text{Integers}
\end{array}
$$

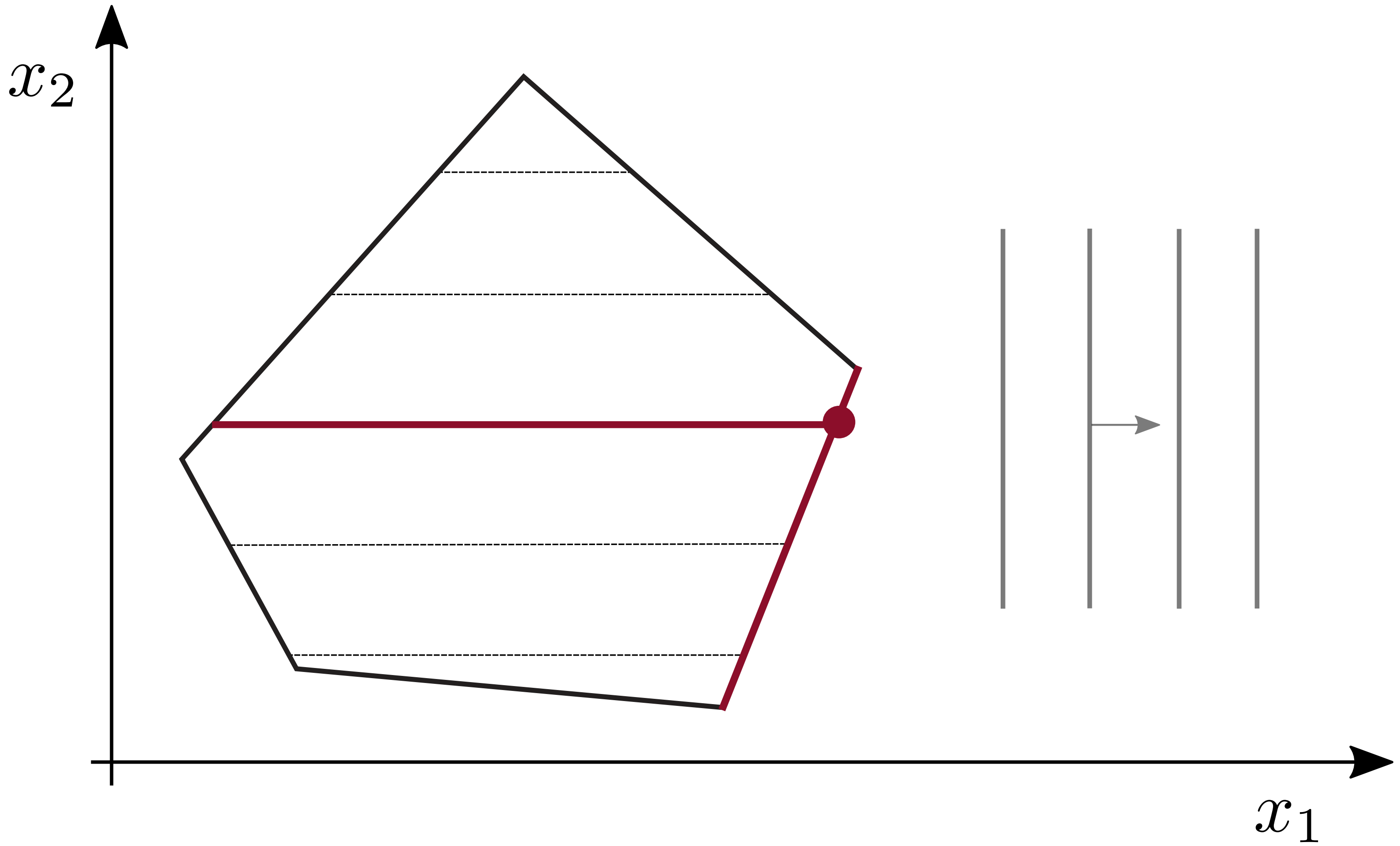
Tight constraints are not enough



Strategy

$$s(\theta) = (\mathcal{T}(\theta), x^\star_{\mathcal{I}}(\theta))$$

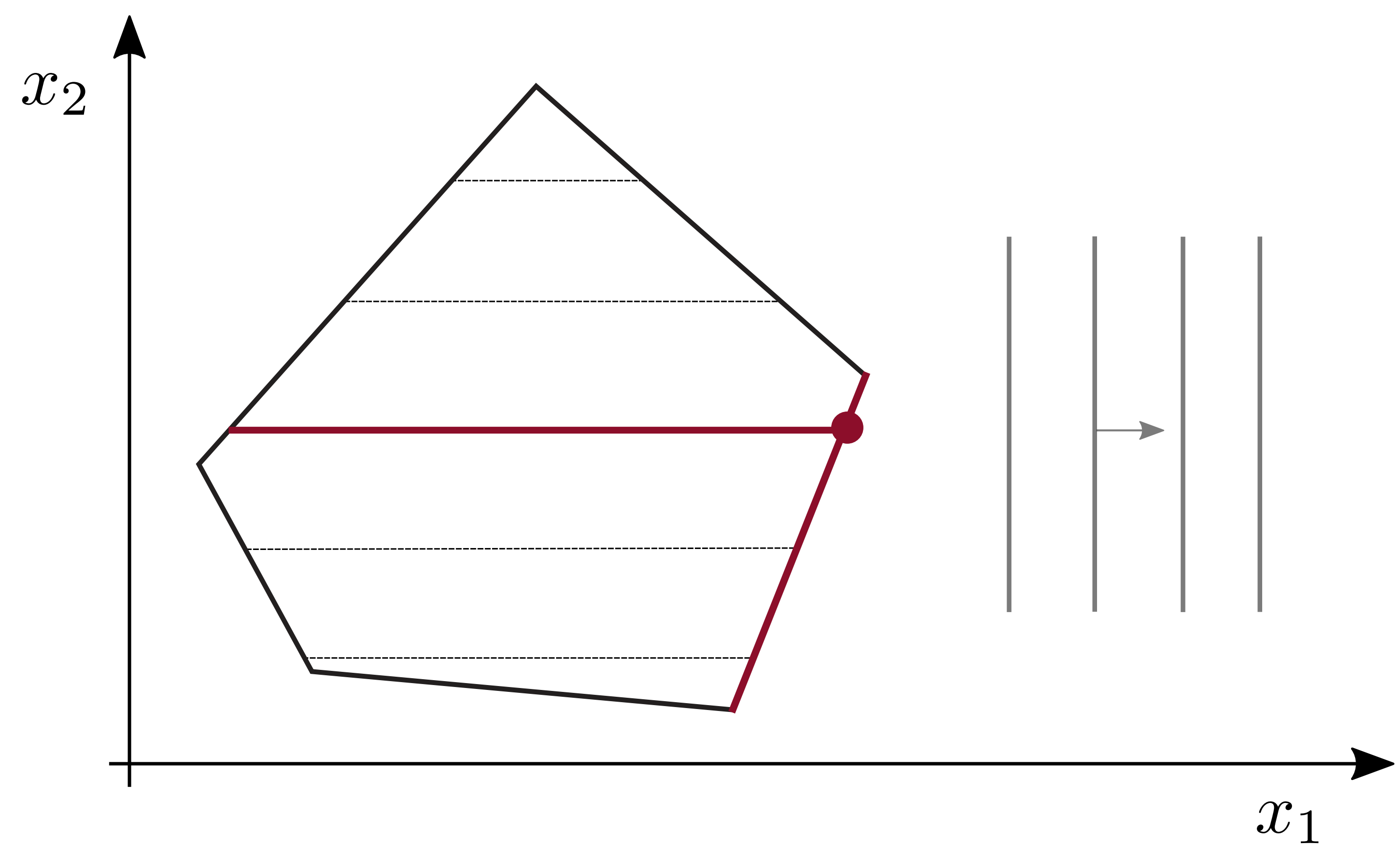
Integer variables



$$s(\theta) \longrightarrow \boxed{\text{Solution Decoding}} \longrightarrow x^\star$$

$$\begin{array}{ll} \text{minimize} & c(\theta)^T x \\ \text{subject to} & A_i(\theta)x = b_i(\theta) \quad \forall i \in \mathcal{T}(\theta) \\ & \boxed{x_\mathcal{I} = x_\mathcal{I}^\star(\theta)} \end{array}$$

Integers

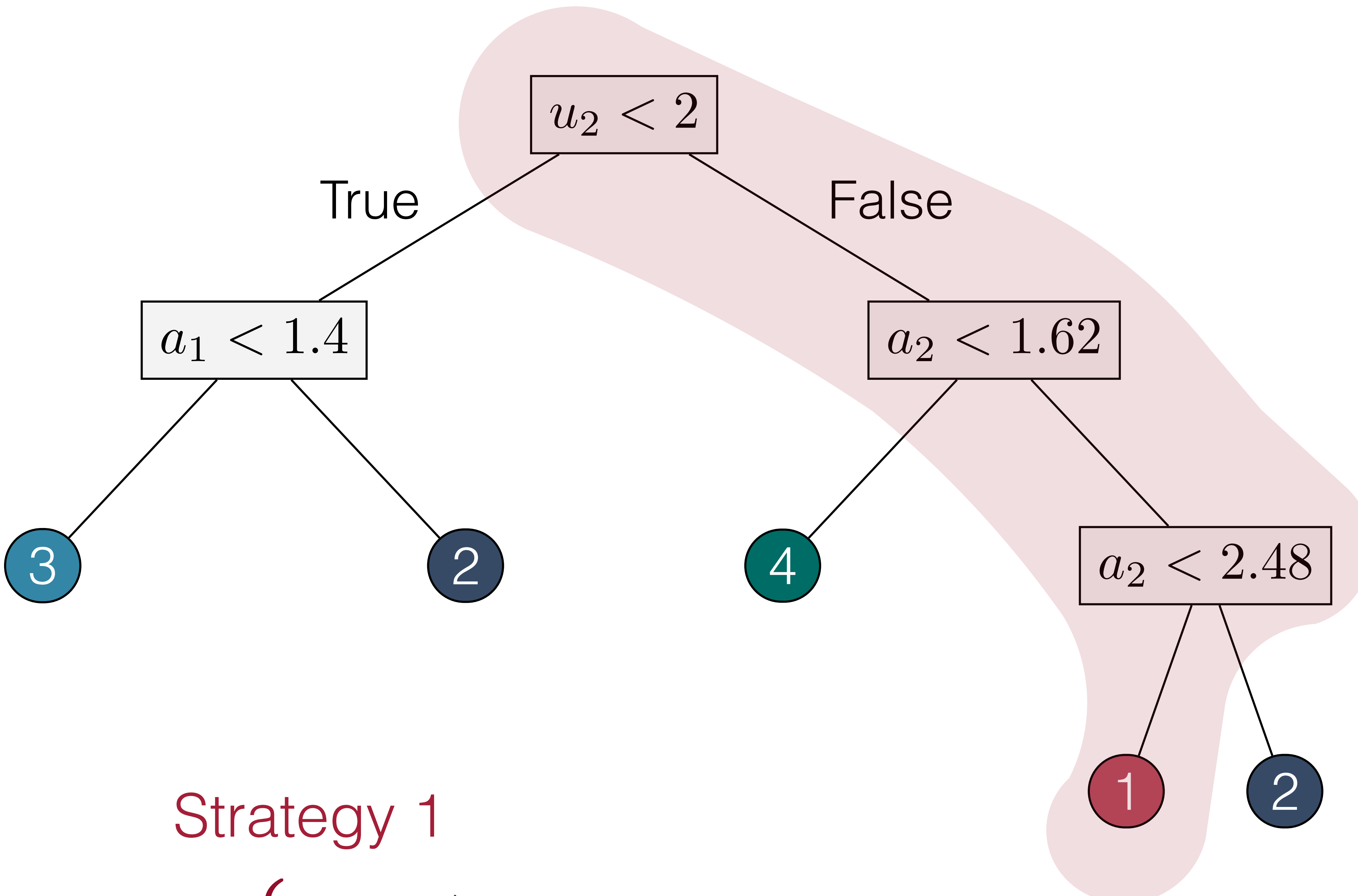


Knapsack

$$\begin{array}{ll} \text{maximize} & c^T x \\ \text{subject to} & \boxed{a}^T x \le b \\ & 0 \le x \le \boxed{u} \\ & x \in \mathbf{Z}^n \end{array}$$

Parameters

Strategy selection

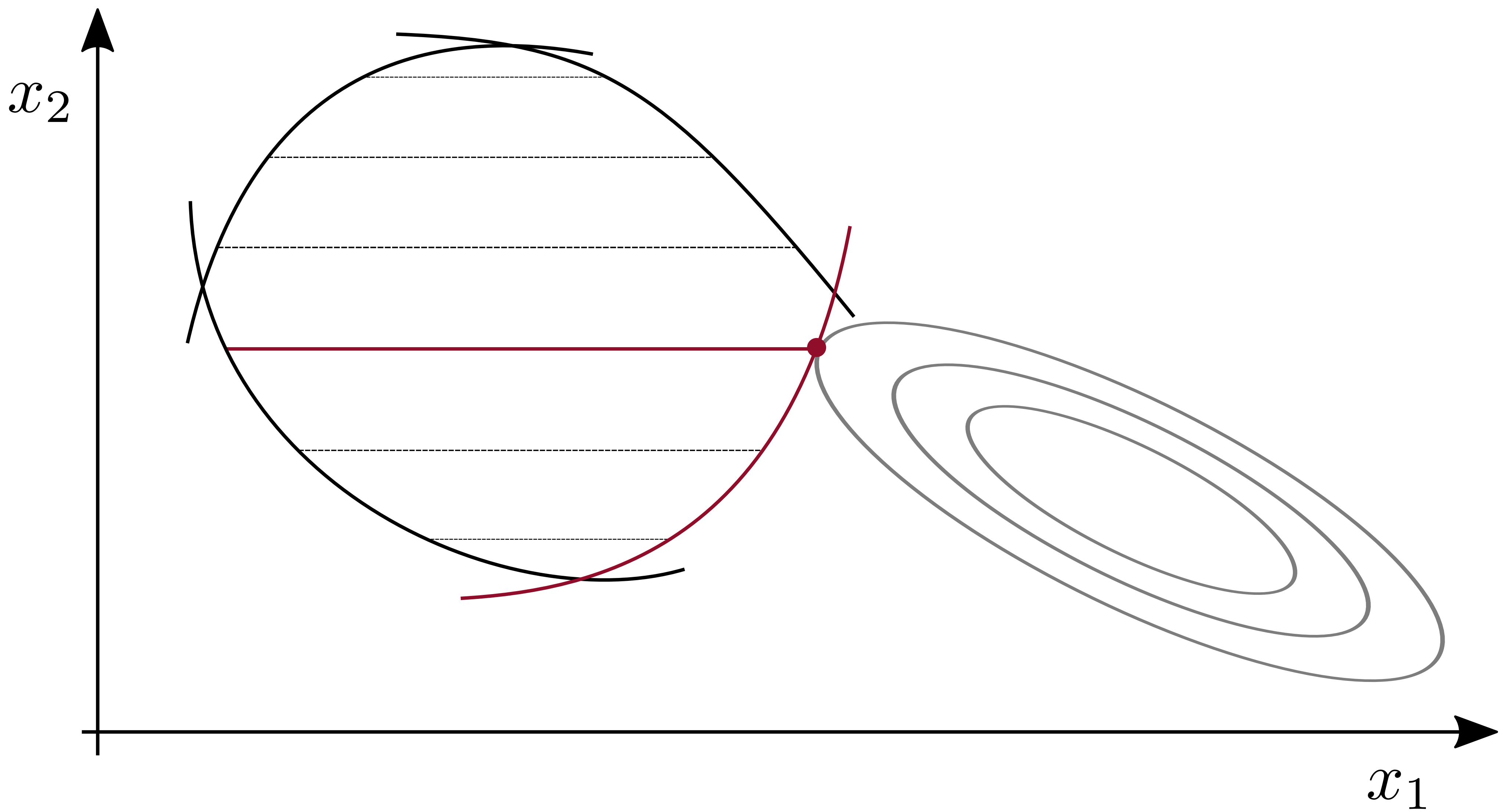


Strategy 1

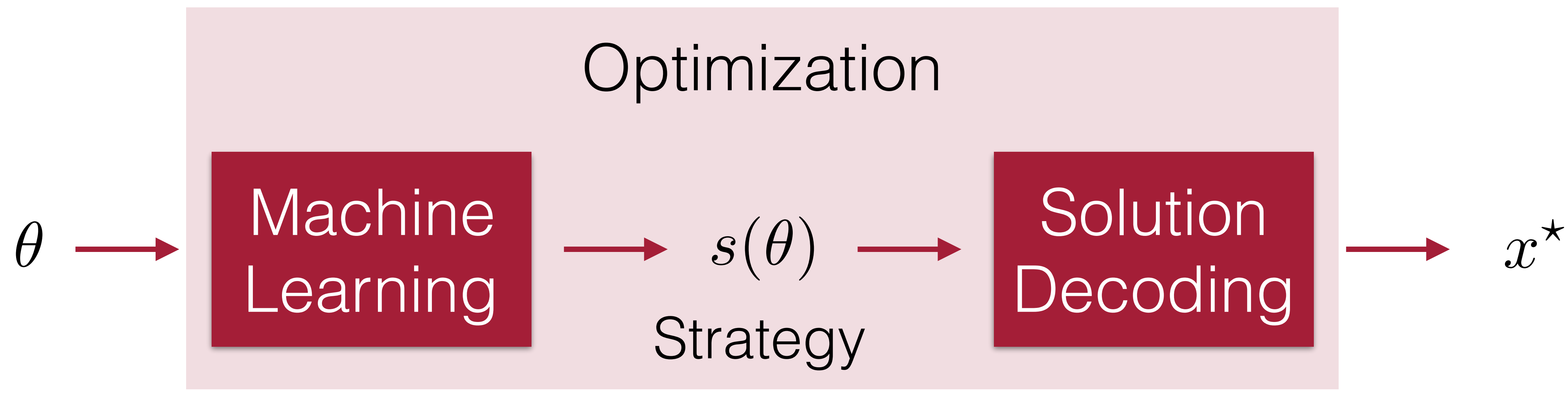
$$x_i = \begin{cases} 0 & i \neq 2 \\ 2 & i = 2 \end{cases}$$

Mixed-integer convex optimization

$$
\begin{array}{ll}
\text{minimize} & f(\theta, x) \\
\text{subject to} & g(\theta, x) \leq 0 \\
& x_{\mathcal{I}} \in \mathbf{Z}^d
\end{array}
$$



Learning the strategies



Black box → Open the box

Too slow → Solve in milliseconds

Classification problem

N data

$(\theta_i, s(\theta_i))$

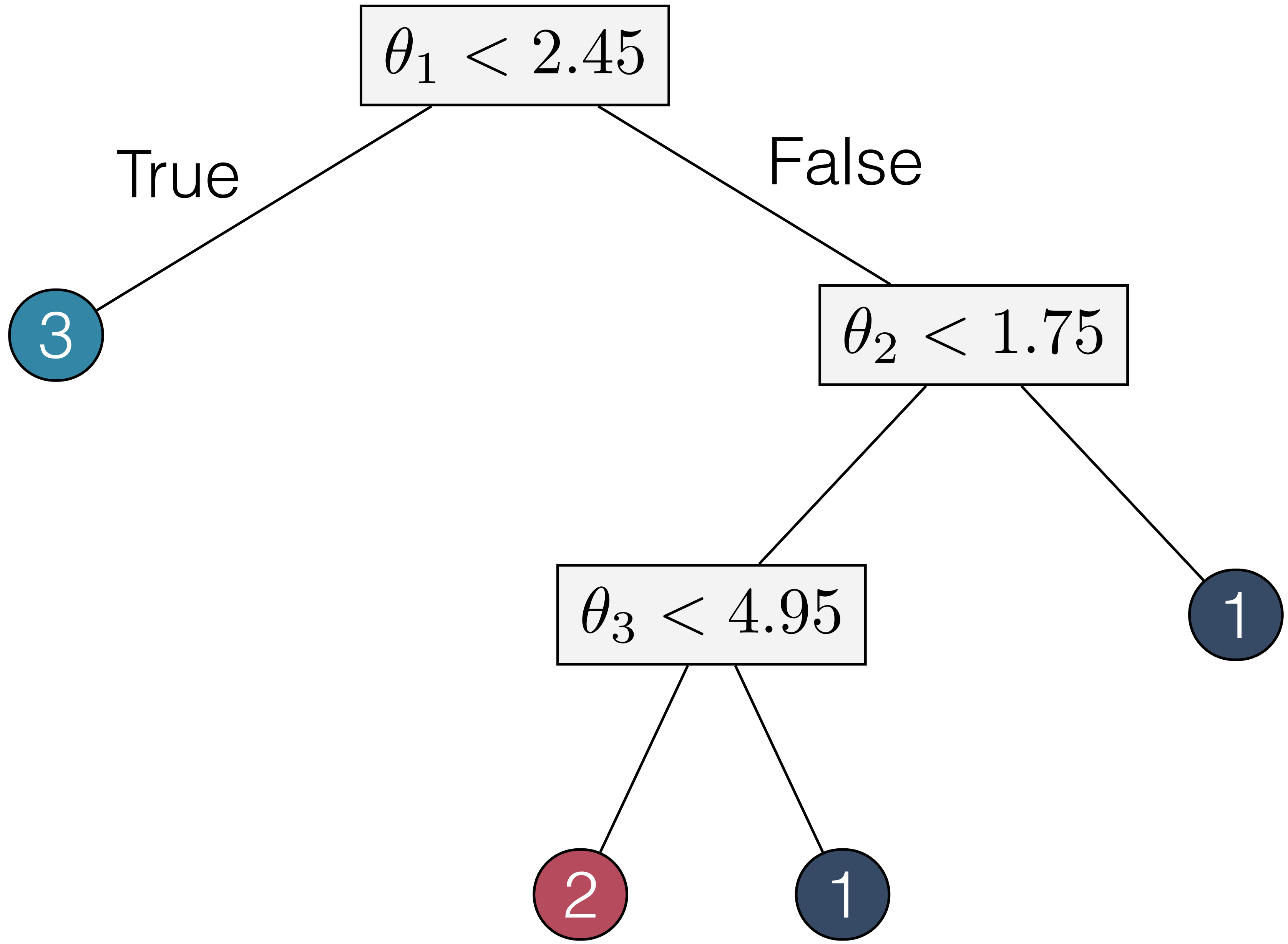
M labels

Strategies $\mathcal{S}$

Learn multiclass classifier

$$\theta \longrightarrow \boxed{\text{Machine Learning}} \longrightarrow \hat{s}(\theta)$$

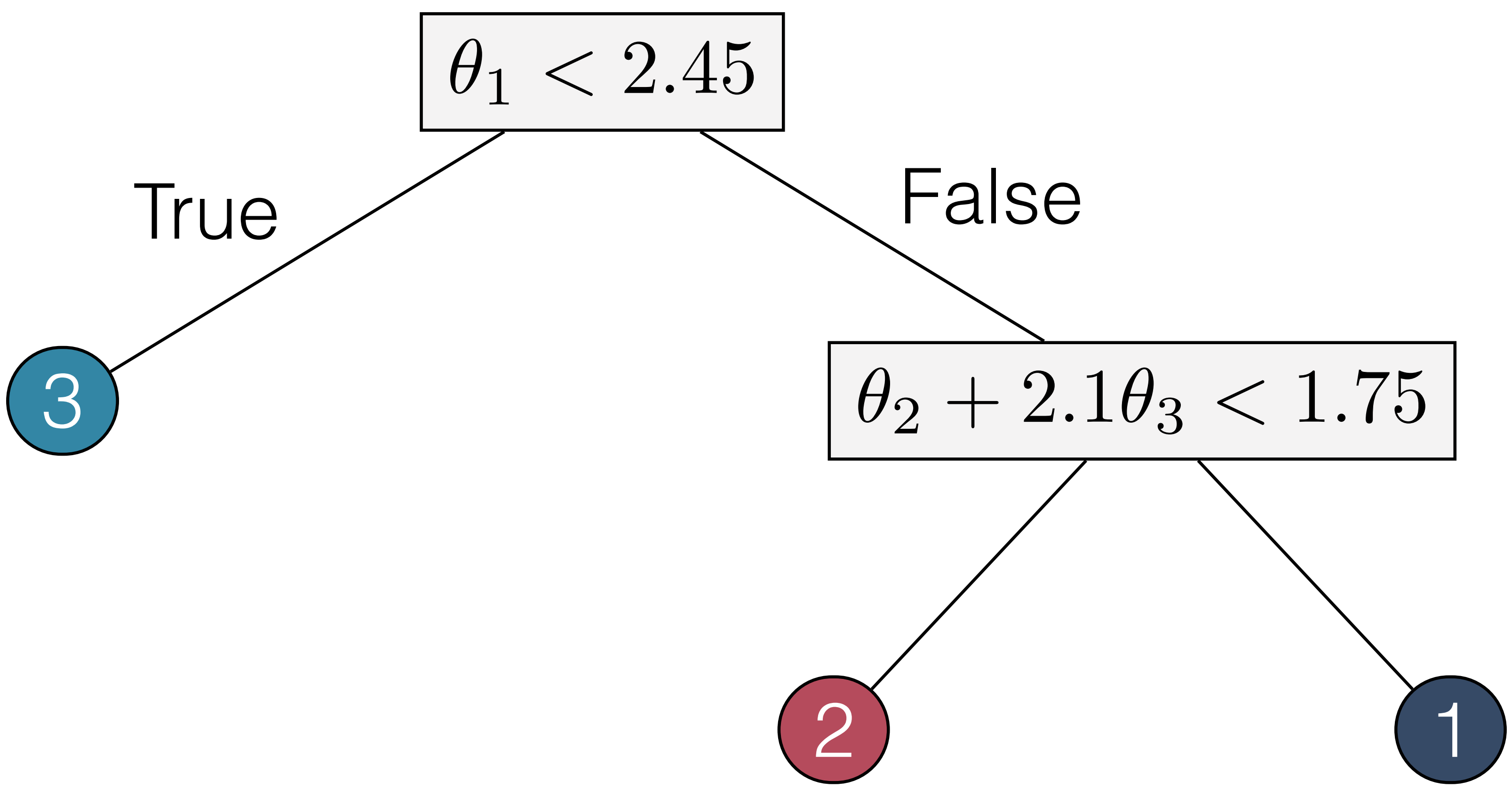
Optimal Classification Trees (OCT)



Interpretable AI

[Bertsimas and Dunn (2017)]

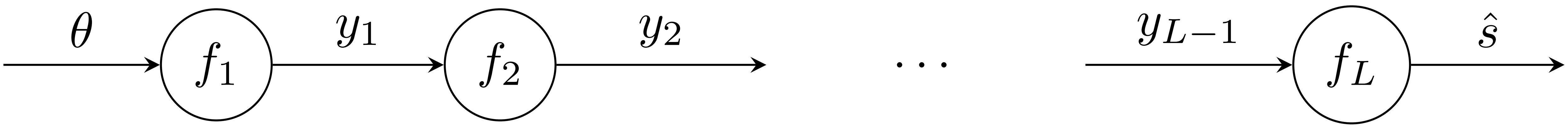
Optimal Classification Trees with Hyperplanes (OCT-H)



Interpretable AI

[Bertsimas and Dunn (2017)]

Neural Networks (NN)



Single layer

$$y_l = f(y_{l-1}) = (W_l y_{l-1} + b_l)_+$$

PyTorch

Strategies exploration

Classification problem

N data

$(\theta_i, s(\theta_i))$

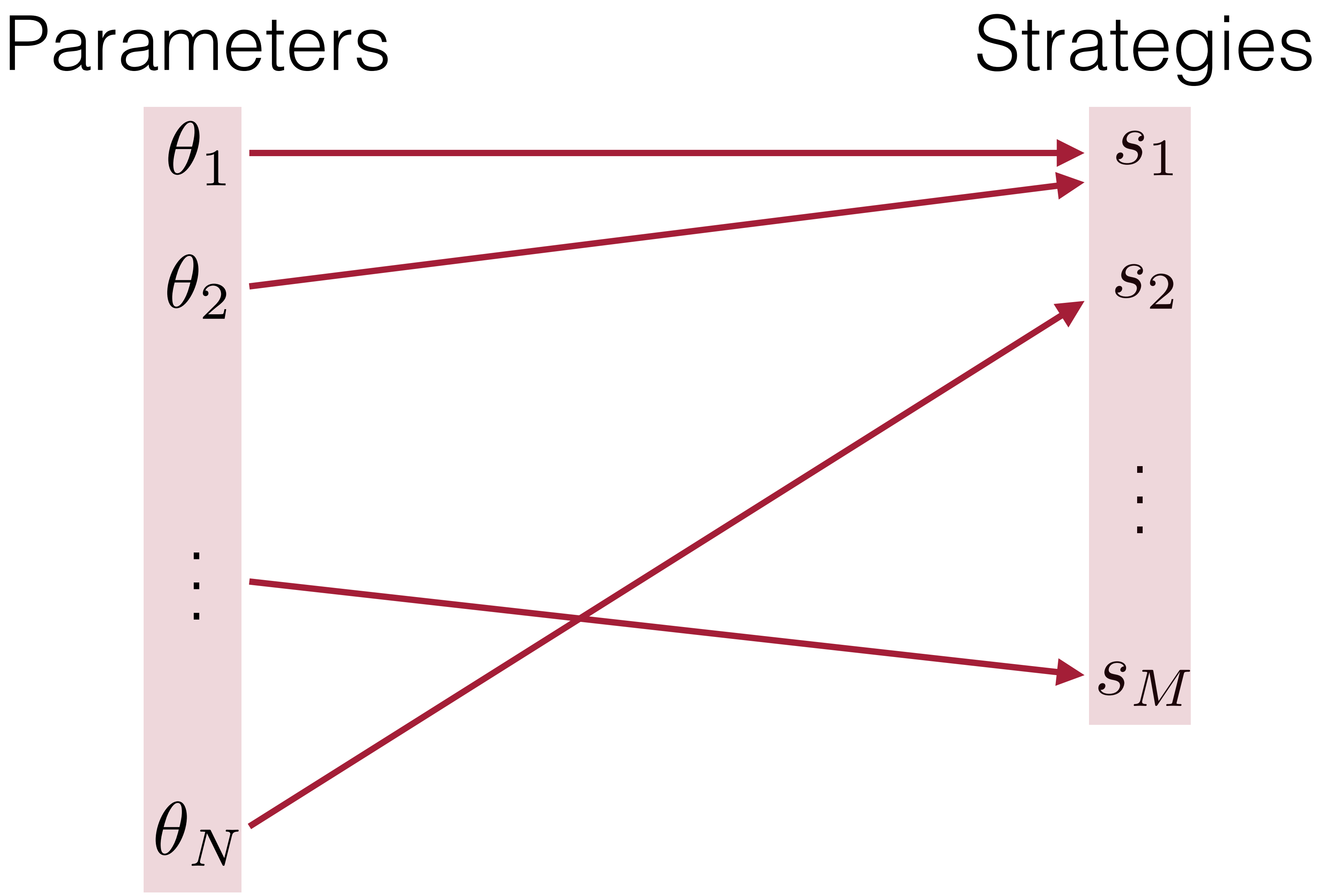
M labels

Strategies $\mathcal{S}$

$\theta \longrightarrow$ **Machine Learning** $\longrightarrow \hat{s}(\theta)$

Have we seen enough data?

Will we find new strategies?



θ_{N+1}?

Alan Turing already knew that

Decoded the **Enigma Machine**
in
World War II

Only some words (labels) needed

Good-Turing estimator

s_1	12 times
s_2	45 times
$\vdots$	$\vdots$
s_M	2 times

$GT \approx$ probability of unseen strategies

$$GT = \frac{N_1}{N}$$

N_1: # strategies appeared once

N: # samples

Sampling scheme

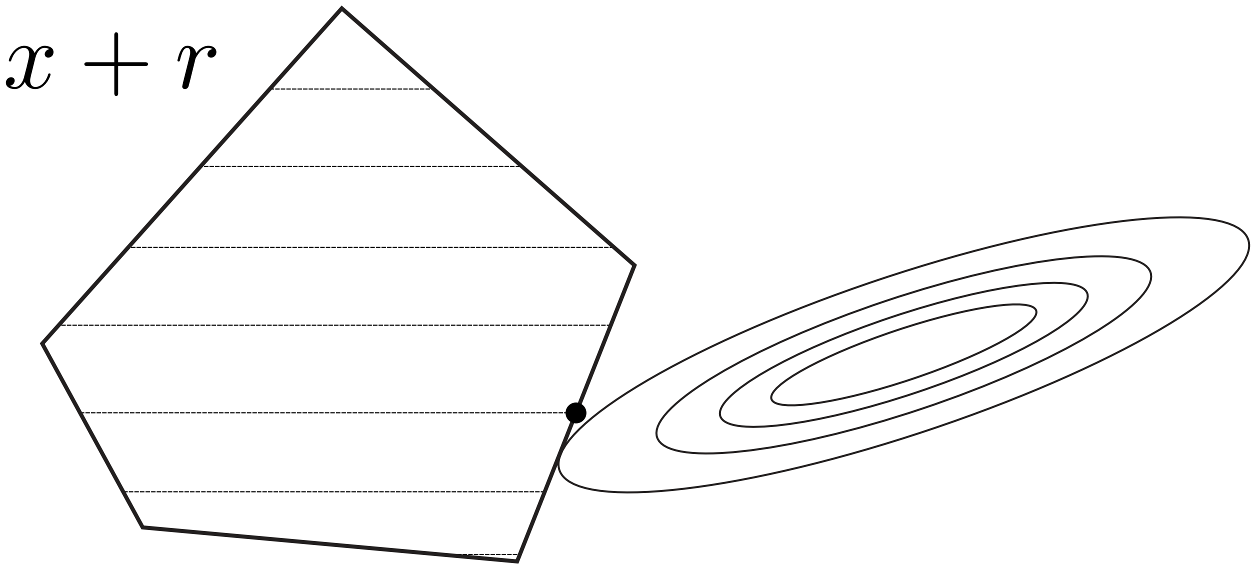
Repeat until $GT \leq \epsilon$

1. sample θ_i
2. compute $s(\theta_i)$
3. update estimator $GT \leftarrow \dfrac{N_1}{N}$

Speedups

Mixed-integer quadratic optimization

$$\begin{array}{ll} \text{minmimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Ax \leq b \\ & x_{\mathcal{I}} \in \mathbf{Z}^d \end{array}$$



Strategy

$$\begin{array}{ll} \text{minmimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & A_{\mathcal{T}(\theta)} x = b_{\mathcal{T}(\theta)} \\ & x_{\mathcal{I}} = x^\star_{\mathcal{I}}(\theta) \end{array}$$

Equality Constrained QP

Quick solution

KKT System

$$\begin{bmatrix} P & A_{\mathcal{T}(\theta)}^T & I_{\mathcal{I}}^T \\ A_{\mathcal{T}(\theta)} & & \\ I_{\mathcal{I}} & & 0 \end{bmatrix} \begin{bmatrix} x \\ \nu \end{bmatrix} = \begin{bmatrix} -q \\ b_{\mathcal{T}(\theta)} \\ x_{\mathcal{I}}^{\star}(\theta) \end{bmatrix}$$

$\mathcal{NP}$-hard problem → Linear System

MLOPT

Offline Learning

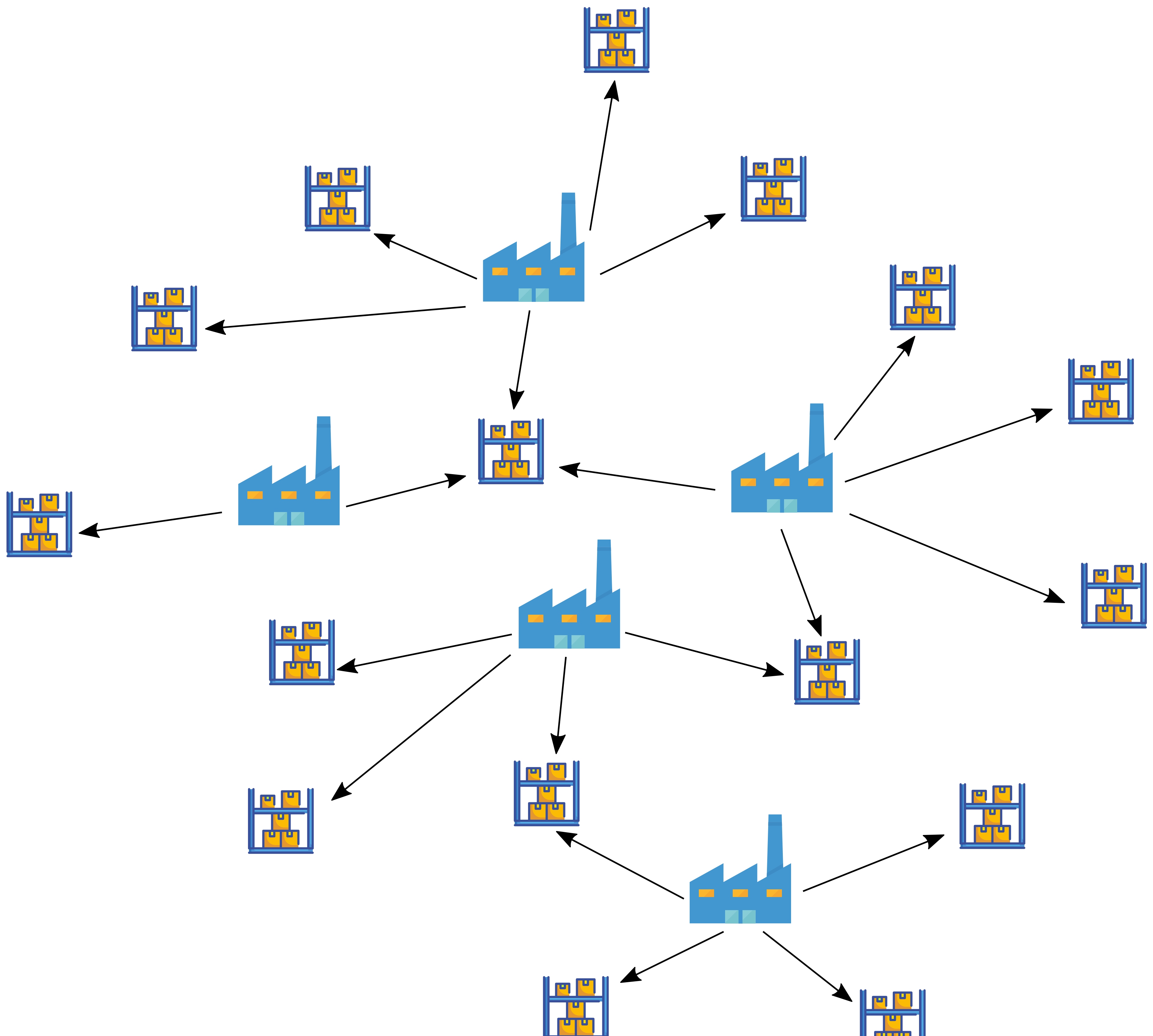
CVXPY modeling $\longrightarrow$ Strategies Sampling $\longrightarrow$ NN/OCT Training

Online Optimization

θ $\longrightarrow$ Strategy Prediction $\longrightarrow$ $s(\theta)$ $\longrightarrow$ Solution Decoding $\longrightarrow$ $x^\star$

Benchmarks

Facility location



Facility location

Shipment cost

Facility cost

minmimize $$\sum_{i \in F} \sum_{j \in W} c_{ij} x_{ij} + \sum_{i \in I} f_i y_i$$

subject to

$$\sum_{i \in F} x_{ij} \geq d_j \quad \forall j \in W$$

Demand

$$\sum_{j \in W} x_{ij} \leq s_i y_i, \quad \forall i \in F$$

Supply

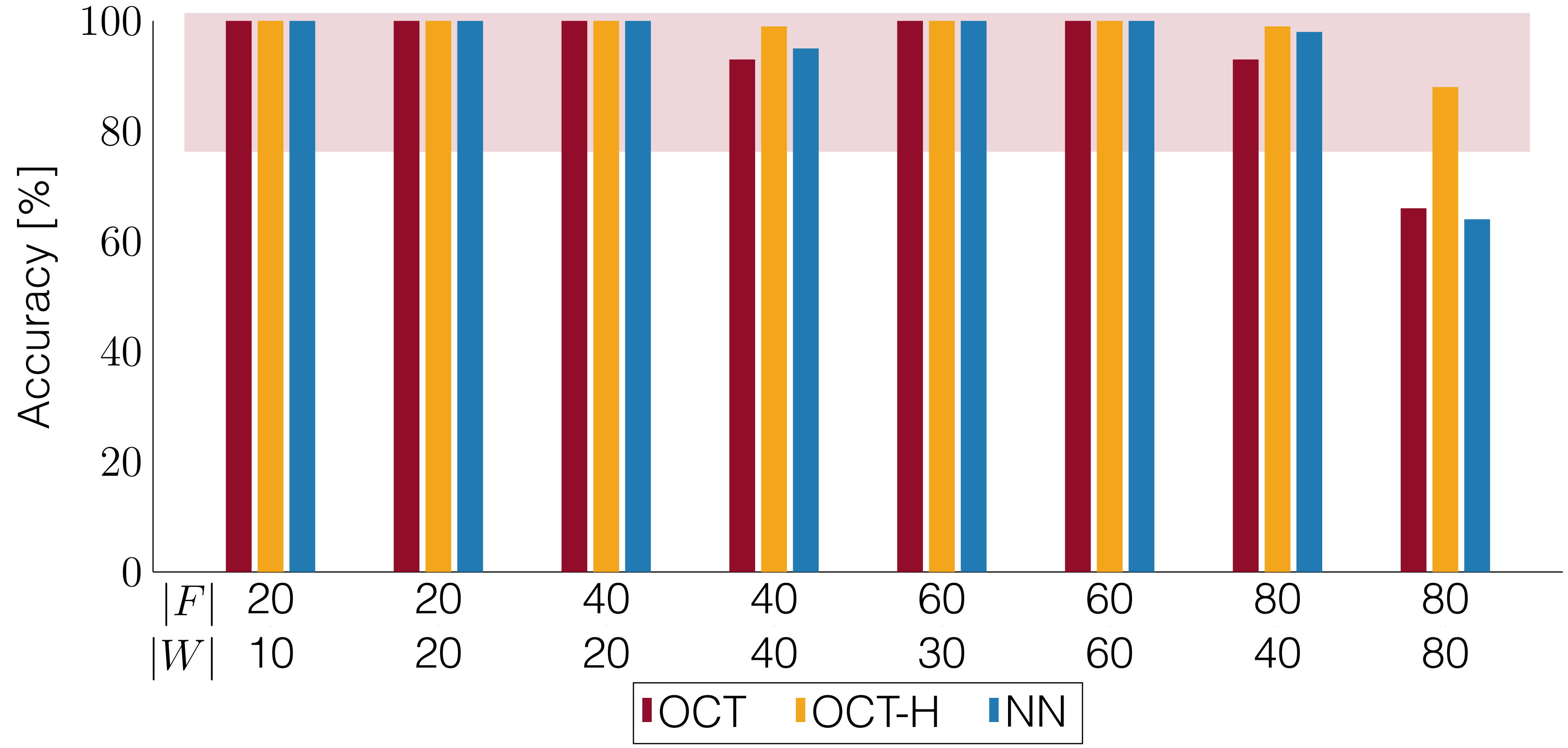
Parameters

$$x_{ij} \geq 0, y_i \in \{0, 1\}$$

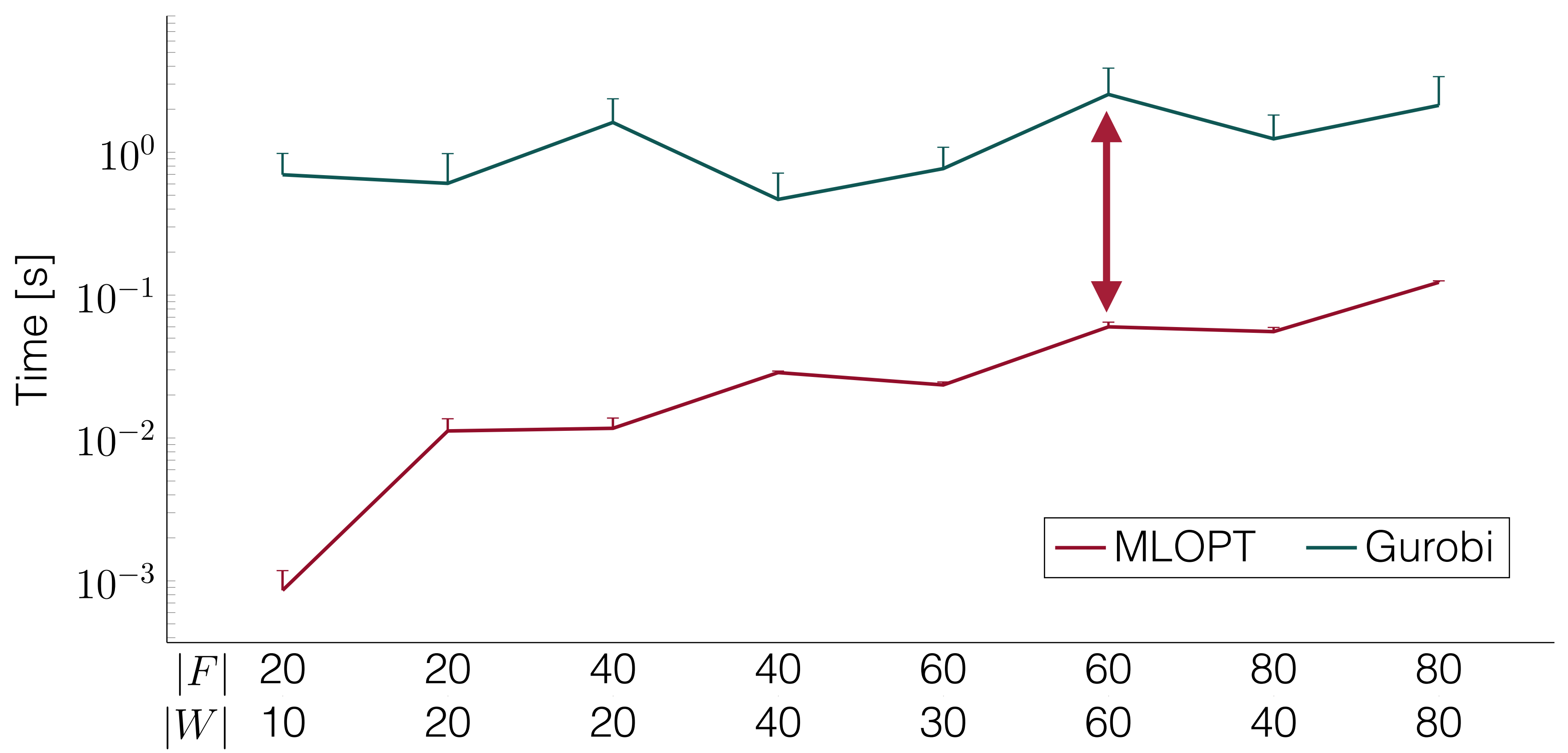
Shipment

Facility

Facility location results

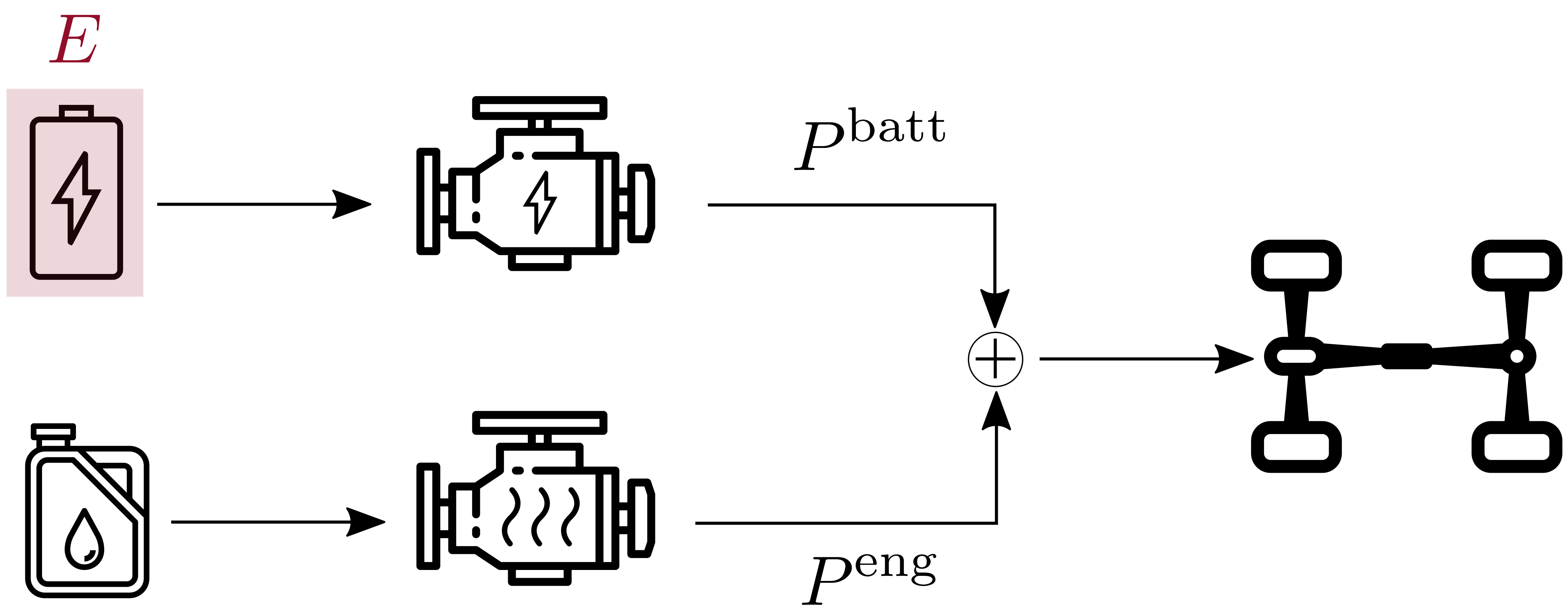


High accuracy



100x faster

Hybrid-vehicle control



Hybrid-vehicle control

$$\text{maximize} \quad \sum_{t=0}^{T-1} f(P_t^{\text{eng}}, z_t) + \delta(z_t - z_{t-1})_+$$

$$\text{subject to} \quad P_t^{\text{batt}} + P_t^{\text{eng}} \geq \boxed{P_t^{\text{des}}}$$

Desired Power

$$E_{t+1} = E_t - \tau P_t^{\text{batt}}$$
$$E_0 = \boxed{E_{\text{init}}}$$

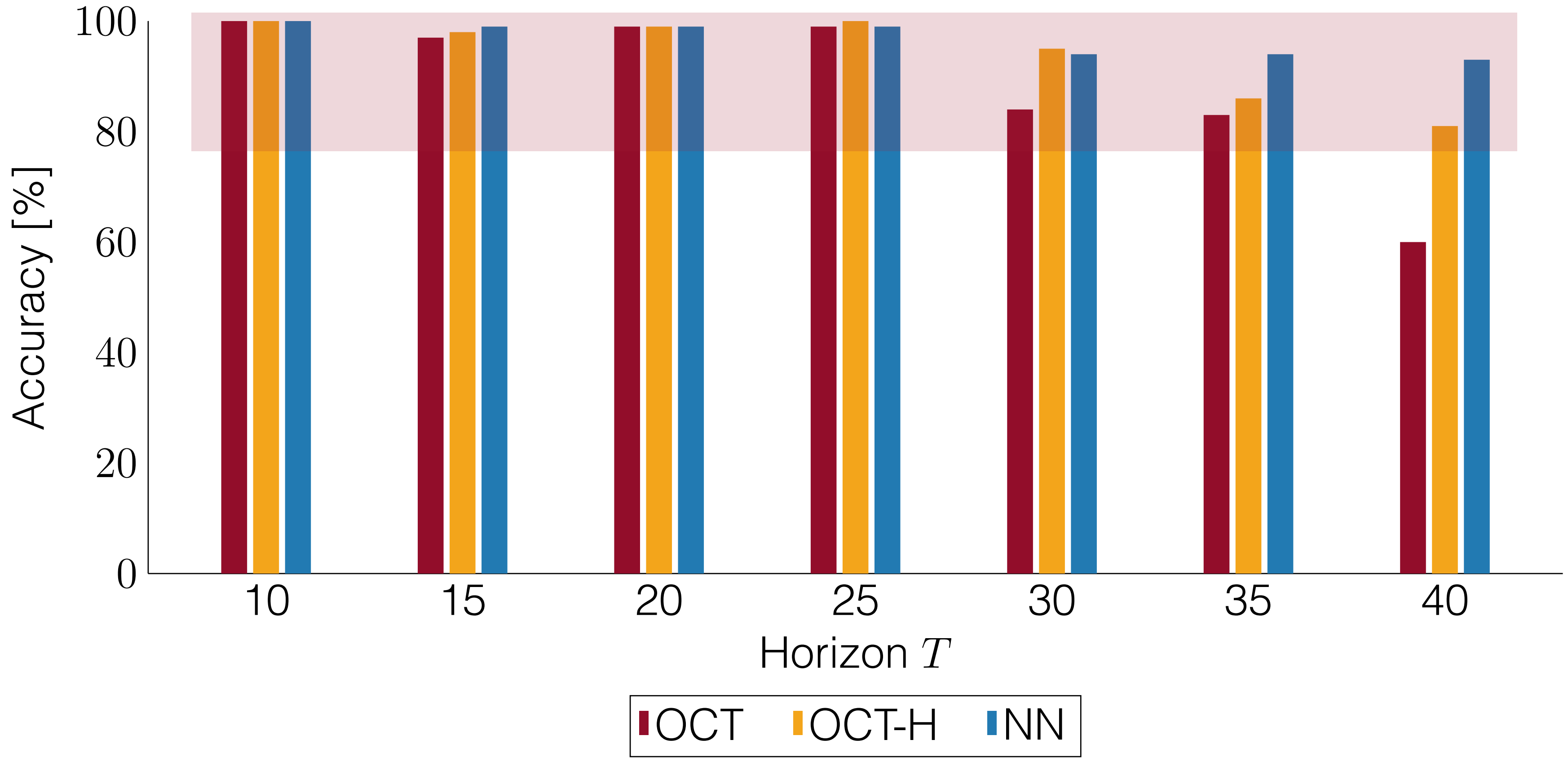
Battery Dynamics

$$z_t \in \{0, 1\}$$

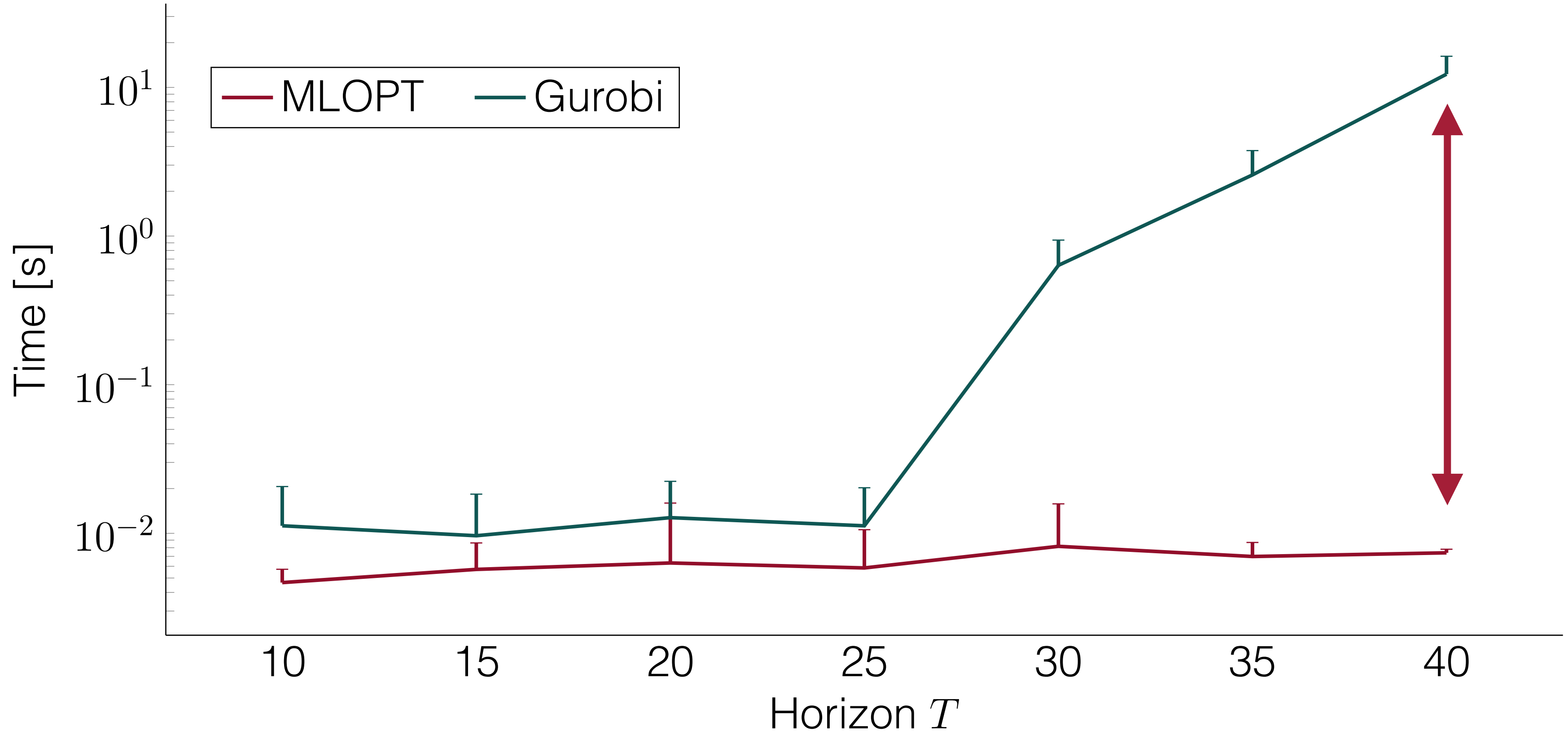
Engine Switch

Parameters

Model-predictive control results

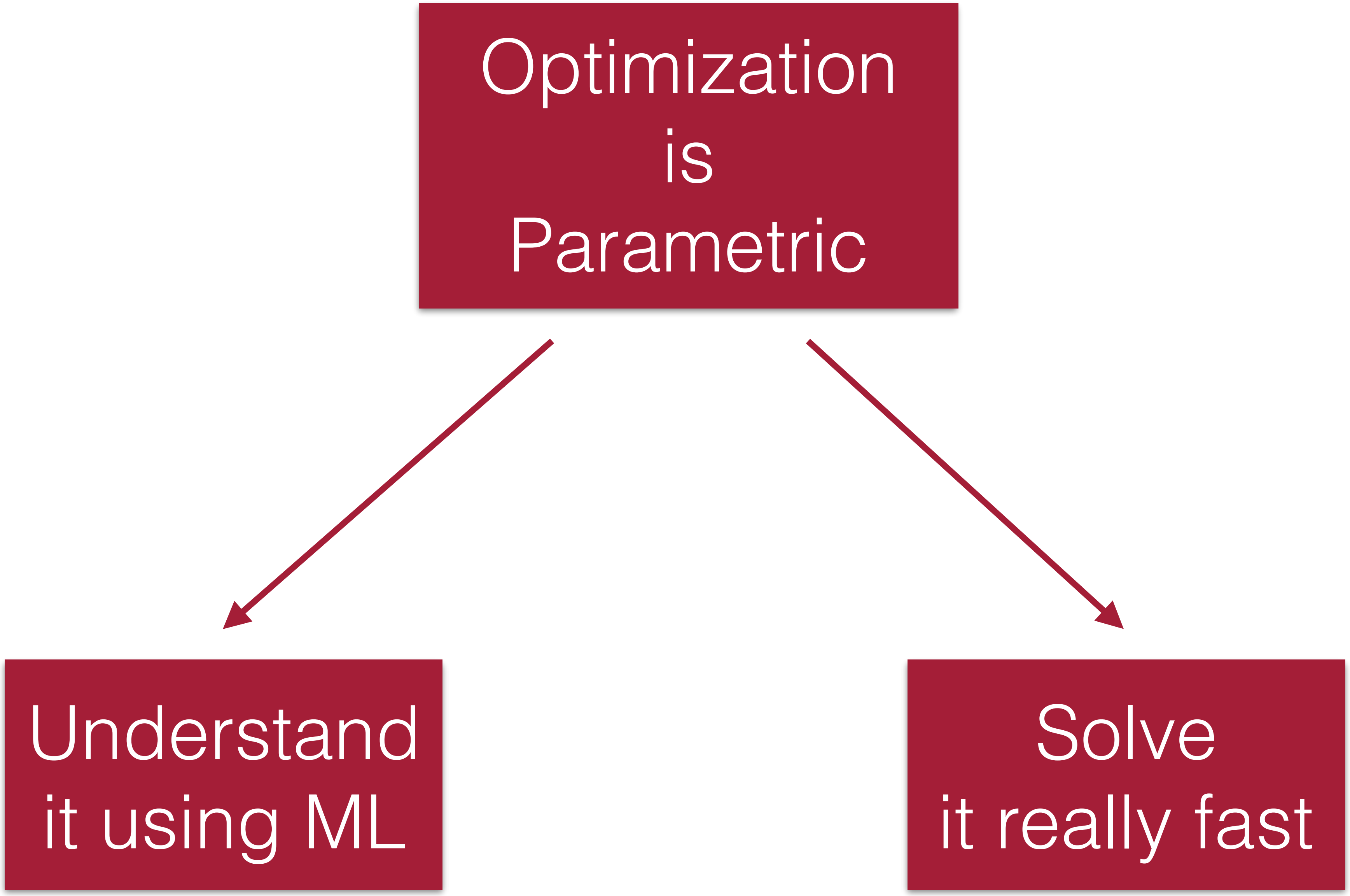


High accuracy



1000x faster

Conclusions



Bertsimas, D., Stellato, B., *The Voice of Optimization*, arXiv:1812.09991, 2018

Backup

Good-Turing estimator

Approximation

$$GT = \frac{N_1}{N} \approx \mathbf{P}(\theta_{N+1} \in \mathbf{R}^p \mid s(\theta_{N+1}) \notin \mathcal{S}(\Theta_N))$$

Bound

$$\mathbf{P}(\theta_{N+1} \in \mathbf{R}^p \mid s(\theta_{N+1}) \notin \mathcal{S}(\Theta_N)) \leq GT + c\sqrt{(1/N)\ln(3/\beta)}$$

Sampling scheme

Algorithm 1 Strategies exploration

given $\epsilon, \beta, \Theta = \emptyset, \mathcal{S} = \emptyset, u = \infty$
for $k = 1, \ldots,$ **do**
  Sample θ_k and compute $s(\theta_k)$ ▷ Sample parameter and strategy.
  $\Theta \leftarrow \Theta \cup \{\theta_k\}$ ▷ Update set of samples.
  if $s(\theta_k) \notin \mathcal{S}$ **then**
    $\mathcal{S} \leftarrow \mathcal{S} \cup \{s(\theta_k)\}$ ▷ Update strategy set if new strategy found
  end if
  if $G + c\sqrt{(1/k)\ln(3/\beta)} \leq \epsilon$ **then** ▷ Break if bound less than ϵ
    break
  end if
end for
return $k, \Theta, \mathcal{S}$